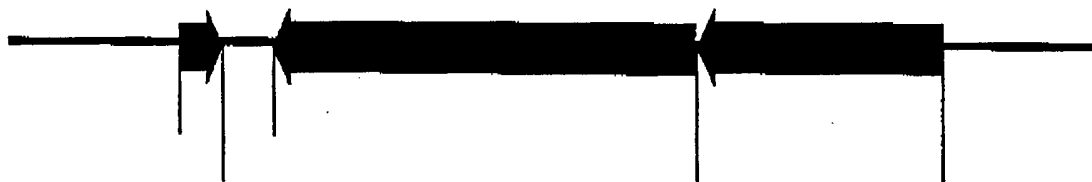


**Figure 1**

Schematic layout of the arrangement of the genetic locus encoding the signal peptide precursor, the histidine kinase and the response regulator. Note that this arrangement is different from other loci in related streptococci for the following reasons: a) The *comC* gene is transcribed from its own promoter alone, unlike the genes thus far described in other streptococci that are arranged in an operon-like cluster with the *comC/DE* genes being transcribed from a single promoter.

b) The *comC* gene is separated from the *comD* gene by 148 nucleotides.

**Streptococcus mutans  
*ComCDE* Operon**



09633017-041001  
T00T40-2T0E860

Figure 2

Sequences of the open reading frames encoding the signal peptide precursor (ComC), the histidine kinase (ComD), and the response regulator (ComE).

> *S. mutans* comC gene

Encodes a precursor to a signal peptide

[ATGAAAAAACACTATCATTAAAAAATGACTTTAAAGAAATTAAGACTGATGAATTAG  
AGATTATCATTGGCGGA (AGCGGAAGCCTATCAACATTTTTCCGGCTGTTTAACAGAAG  
TTTTACACAAGCTTTGGGAAAA)] TAA

> *S. mutans* CSP encoding sequence

Competence Signal Peptide

AGCGGAAGCCTATCAACATTTTTCCGGCTGTTTAACAGAAGTTTACACAAGCTTTGGG  
AAAA [SEQ ID NO:1]

> *S. mutans* comD gene

Encodes a protein that functions as a histidine kinase  
receptor

[ATGAATGAAGCCTTAATGATACTTTCAAATGGTTTATTAAGTTATCTAACCCTTCTAT  
TTCTCTTGTTTCTATTTTCTAAGGTAAGTAATGTCACTTTATCGAAAAAGGAATTAAGT  
CTTTTTTCGATAAGCAATTTTCTGATAATGATTGCTGTTACGATGGTGAACGTAAACCT  
GTTTTATCCTGCAGAGCCTCTTTATTTTATAGCTTTATCAATTTATCTTAATAGACAGA  
ATAGTCTTTCTCTAAATATATTTTATGGTCTGCTGCCTGTTGCCAGTTCTGACTTGTTT  
AGGCGGGCAATCATATTTCTTTATCTTGATGGAAGTCAAGGAATTGTAATGGGCAGTAG  
CATTATAACCACCTATATGATCGAGTTTGCAGGAATAGCGCTAAGTTACCTCTTTCTCA  
GTGTGTTCAATGTTGATATTGGTCGACTTAAAGATAGTTTGACCAAGATGAAGGTCAAA  
AAACGCTTGATTCCAATGAATATTACTATGCTTCTATACTACCTTTTAATACAGGTATT  
GTATGTTATAGAGAGTTATAATGTGATACCGACTTTAAATTTTCGTAAATTTGTCGTTA  
TTGTCTATCTTATTTTATTTTGTATTCTGATCTCATTTTTAAGCCAATATACCAAACAA  
AAGGTTCAAATGAGATAATGGCACAAGGAAGCTCAGATTGCAATATACCCAGTA  
TAGTCAGCAAATAGAATCTCTTACAAGGATATTGCAAGTTTCCGCCATGATTATCTGA  
ATATTTTAACTAGCCTCAGATTAGGCATTGAAAATAAAGATTTAGCTAGTATTGAAAAG  
ATTTACCATCAAATCTTAGAAAAAACAGGACATCAATTGCAGGATACCCGTTATAATAT  
CGGCCATCTAGCTAATATTCAAAACGATGCTGTCAAGGGTATCTTGTCAGCAAAATCT  
TAGAAGCTCAGAATAAAAAGATTGCTGTCAATGTAGAAGTCTCAAGTAAAATACAACTG  
CCTGAGATGGAGTTGCTTGATTTCATTACCATACTTTCTATCTTGTTGTGATAATGCCAT  
TGAGGCTGCTTTTGAATCATTAATCCTGAAATTCAGTTAGCCTTTTTTAAGAAAAATG  
GCAGTATAGTCTTTATCATTGAGAATTCACCAAGAAAAACAAATAGATGTGAGTAAA  
ATTTTTAAAGAAAACTATTCCACTAAAGGCTCCAATCGCGGTATTGGTTTAGCAAAGGT  
GAATCATATTTCTTGAACATTATCCCAAACAGTTTACAAACAAGCAATCATCATCATT  
TATTCAGCAACTCCTAATAATAAAA] TAG

> *S. mutans* comE gene

Encodes a response regulator that activates transcription  
of a number of genes

09833017-041001

[ATGATTTCTATTTTTGTATTGGAAGATGATTTTTTACAACAAGGACGTCTTGAAACCA  
CCATTGCAGCTATCATGAAAGAAAAAATTGGTCTTATAAGAATTGACTATTTTTTGA  
AAACCACAACAACCTTATTGACGCTATCCCTGAAAAGGGCAATCACCAGATTTTCTTTT  
GGATATTGAAATCAAAAAAGAGGAAAAGAAAGGACTGGAAGTAGCCAATCAGATTAGAC  
AGCATAATCCTAGTGCAGTTATTGTCTTTGTACAGACACATTCTGAGTTTATGCCCCCTC  
ACTTTTCAGTATCAGGTATCTGCTTTGGATTTTATTGATAAATCTTTGAATCCTGAGGA  
GTTCTCCACCCGATTGAATCAGCGCTGTATTATGCTATGGAAAACAGCCAGAAGAATG  
GTCAATCAGAGGAACTTTTATTTTCCATTCATCTGAAACTCAGTTTCAGGTCCCTTTT  
GCTGAGATTCTGTATTTTGAACATCTTCAACAGCCCATAAGCTCTGCCTTTATACTTA  
TGATGAACGGATTGAATTCTACGGCAGTATGACTGACATTGTTAAAATGGATAAGAGAC  
TTTTTCAGTGCCATCGCTCTTTTATTGTCAATCCTGCCAATATTACCCGTATTGATCGG  
AAAAAACGCTTGCCCTATTTTCGAAATAATAAGTCTTGTCTTATTTACGAACTAAGTT  
AACAAAACCTGAGAGCTGTGATTGCTGATCAAAGGAGAGCAAAA] TGA

09833017-041001

The amino acid sequences of the signal peptide precursor (ComC), the histidine kinase (ComD), and the response regulator (ComE).

MRKTLSLKNDFKELKTDELEIIIGGSGSLSTFFRLFNRSFTQALGK

MNEALMILSNGLLTLYLTVLFLFLFLFSKVSNTLSKKELTFSISNFLIMIAVTMVNVNL  
FYPAEPLYFIALSIYLNQRQNSLSLNIIFYGLLPVASSDLFRRAIFFILDGTQGIVMGSS  
IITTYMIEFAGIALSYLFLSVFNVDIGRLKDSLTKMKVKKRLIPMNITMLLYYLLIQVL  
YVIESYNVIPTLKFRKFVIVYLILFLILISFLSQYTKQKVQNEIMAOKEAQIRNITQY  
SQQIESLYKDIFSFRHDYLNILTSRLGIENKDLASIEKIYHQILEKTGHQLQDTRYNI  
GHLANIQNDAVKGILSAKILEAQNKKIHAVNVEVSSKIQLPEMBLLDFITILSILCDNAI  
EAAFESLNPEIQLAFFKKNGSIVFIIQNSTKEKQIDVSKIFKENYSTKGSNRGIGLAKV  
NHILEHYPKTSLQTSNHHHLFKQLLIK

MISIFVLEDDFLQQGRLETTIAAIMKEKNWSYKELTIFGKPPQOLIDAIPEKGNHQIFFL  
DIEIKKEEKKGLEVANQIRQHNPSAVIVFVTTTHSEFMPLTFQYQVSALDFIDKSLNPEE  
FSHRIESALYYAMENSQKNGQSEELFIFHSSETQFQVPFAEILYFETSSTAHKLCLYTY  
DERIEFYGSMTDIVKMDKRLFQCHRSFIVNPNANITRIDRKKRLAYFRNNKSCLISRTKL  
TKLRAVIADQRRAK

Figure 4

The deduced amino acid sequence of the signal peptide precursor in various strains and its predicted cleavage site. The original peptide is expressed as a 46-amino acid peptide that is cleaved after the glycine-glycine residues to generate an active signal peptide.

```
BM71 CSP      1 MKKTPSLKNDKFKEIKTDELEIIIGSGSLSTFFRLFNRSFTQALGK 46
GB14 CSP      1 MKKTLCLKNDKFKEIKTDELEIIIGSGSLSTFFRLFNRSFTQALGK 46
H7 CSP        1 MKKTLCLKNDKFKEIKTDELEIIIGSGSLSTFFRLFNRSFTQALGK 46
JH1005 CSP    1 MKKTLCLKNDKFKEIKTDELEIIIGSGTLSTFFRLFNRSFTQA 43
LT11 CSP      1 MKKTLCLKNDKFKEIKTDELEIIIGSGSLSTFFRLFNRSFTQALGK 46
NG8 CSP       1 MKKTLCLKNDKFKEIKTDELEIIIGSGSLSTFFRLFNRSFTQALGK 46
UAB159 CSP    1 MKKTLCLKNDKFKEIKTDELEIIIGSGSLSTFFRLFNRSFTQALGK 46
      **** *****
```

consensus: 1 MKKTLCLKNDKFKEIKTDELEIIIG SGSLSTFFRLFNRSFTQALGK 46  
predicted cleavage site: ^

Figure 5

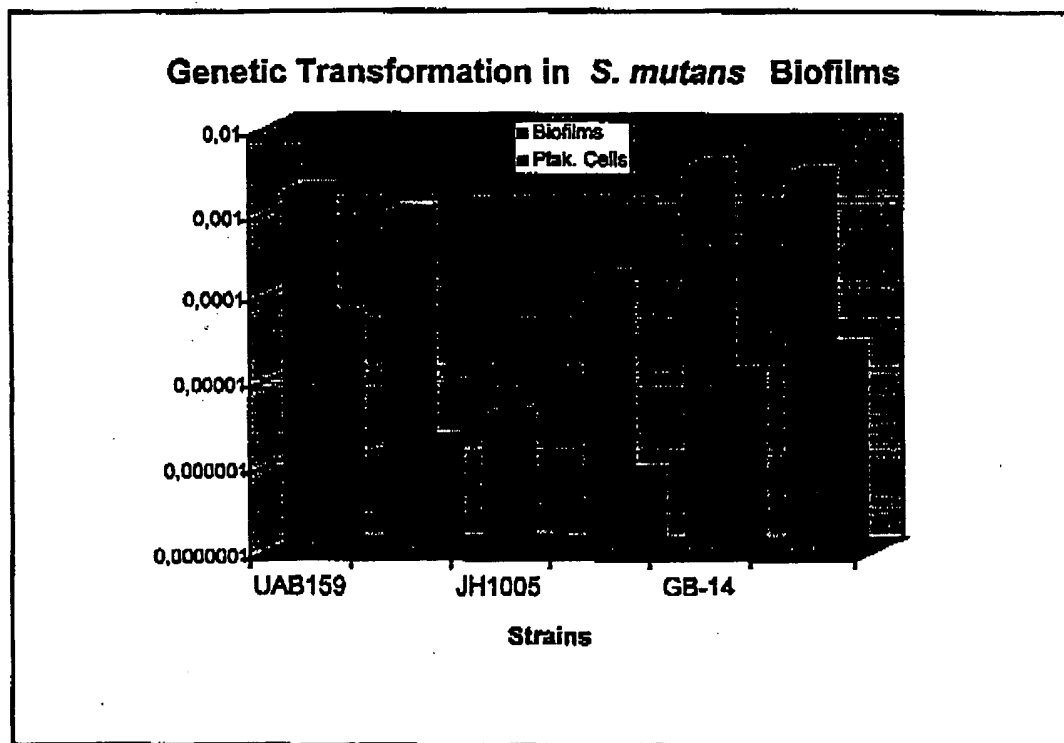
The synthetic signal peptide that is effective at inducing competence, biofilm formation and acid tolerance in *Streptococcus mutans*.

SGSLSTFFRLFNRSFTQALGK [SEQ ID NO:2]

09833017 041001

Figure 6

The natural activity of the signal/receptor system functioning *in vitro* in model biofilms as determined by the ability of various strains of *S. mutans* to accept donor plasmid DNA conferring erythromycin resistance.



09833017-041001

Strain	Peptide added Number of Transformants/Recipients	No peptide Number of Transformants/Recipients
UAB15	$4.65 \times 10^{-1}$	$1.78 \times 10^{-6}$
JH1005 <sup>2</sup>	$6.98 \times 10^{-2}$	0

The final concentration of SCSP used was 500 ng/ml.  
The strain contains a nonsense mutation in the *comC* gene encoding the CSP.

Figure 8

List of the primers used to amplify the genes or internal regions of the target genes by polymerase chain reaction (PCR) for subsequent sequencing or inactivation.

ComC region

ComC Primer Pair: F5-B5

[F5] 23406-23424 5'- AGTTTTTTGTCTGGCTGCG -3'

19 nt forward primer

pct G+C: 47.4 Tm: 50.5

[B5] 24056-24037 5'- TCCACTAAAGGCTCCAATCG -3'

20 nt backward primer

pct G+C: 50.0 Tm: 51.9

651 nt product for F5-B5 pair (23406-24056)

Optimal annealing temp: 50.3

pct G+C: 30.9 Tm: 71.5

ComD region

ComD Primer Pair: F1-B1

[F1] 392-415 5'- CGCTAAGTTACCTCTTTCTCAGTG -3'

24 nt forward primer

pct G+C: 45.8 Tm: 51.6

[B1] 683-663 5'- GCTTCCTTTTGTGCCATTATC -3'

21 nt backward primer

pct G+C: 42.9 Tm: 50.8

292 nt product for F1-B1 pair (392-683)

Optimal annealing temp: 49.5

pct G+C: 30.8 Tm: 70.2

ComE region

ComE Primer Pair: F1-B1

[F1] 145-165 5'- CCTGAAAAGGGCAATCACCAG -3'

21 nt forward primer

pct G+C: 52.4 Tm: 55.9

[B1] 606-585 5'- GCGATGGCACTGAAAAAGTCTC -3'

22 nt backward primer

pct G+C: 50.0 Tm: 55.4

462 nt product for F1-B1 pair (145-606)

Optimal annealing temp: 53.6

pct G+C: 38.3 Tm: 74.1

09833017-041001



Figure 9

ComCDE local region. The ComC (first highlighted region; nucleotides 101 to 241), ComD (second highlighted region; nucleotides 383 to 1708) and ComE (third highlighted region; nucleotides 1705 to 2457) proteins are highlighted.

Sequence Range: 1 to 2557

```

      10      20      30      40      50
ACATTATGTGTCCTAAGGAAAATATTACTTTTTCAAGAAAATCCATGATT
TGTAATACACAGGATTCCTTTTATAATGAAAAGTTCTTTTAGGTACTAA
                                <K K L F I W S K
                                <
      60      70      80      90     100
TTTTCATAAAAATAGTATACTAATTATAATCAAAAAAAGGAGATATAAA
AAAAGTATTTTTTATCATATGATTAATATTAGTTTTTTTCTCTATATTT
<K M F F L I S I I I L F L L Y L
<
     110     120     130     140     150
ATGAAAAAACACTATCATTAAAAATGACTTTAAAGAAATTAAGACTGA
TACTTTTTTTGTGATAGTAATTTTTTACTGAAATTTCTTTAATTCTGACT
M K K K T L S L K N D F A K E I K I D
                                ORF RF [2]
<I F F V S D N F F S K L S I L V S
<
     160     170     180     190     200
TGAATTAGAGATTATCATTGGCGGAAGCGCAAGCCTATCAACATTTTCC
ACTTAATCTCTAATAGTAACCGCCTTCGCCTTCGGATAGTTGTA AAAAGG
B L B I Y I G G S G S E S T F F
                                ORF RF [2]
<S N S I I M
<
     210     220     230     240     250
GGCTGTTTAAACAGAAGTTTTACACAAGCTTTGGGAAAATAAGATAGGCTA
CCGACAAATTGTCTTCAAAATGTGTTGAAACCCCTTTTATTCTATCCGAT
R L F N R S F T O A L G K
                                ORF RF [2]
<
     260     270     280     290     300
ACATTGGAATAAAACAAGGCTGGATTTATTATTCCAGCCTTTTTAAATGT
TGTAACCTTATTTTGTTCGACCTAAATAATAAGGTCCGAAAAATTACA
     310     320     330     340     350
AAAATAAAATACAGGGTTAAATAATCAAGTGTGCTGTCGTGGATGAGAA
TTTTATTTTATGTCCCAATTTATTAGTTCACACGACAGCACCTACTCTT
     360     370     380     390     400
GATAAACTATCTCTTAGAGAATAGGCCTCTCTATTTTATTATTAGGAG
CTATTTTGATAGAGAATCTCTTATCCGAGGAGATAAAATAATAATCCTC
                                <K I L L I
                                < ORF RF [
     410     420     430     440     450
```

09833017 041001

AACGAACTTATTTACTACTACTAACGAACAAACATTTGACCAAACCCCTA  
<O K T L H E E N S T O L S T K T E S

< ORF RF[4] C

460 470 480 490 500  
AATGTTCAAGAATATGATTACCTTTGCTAAACCAATACCGCGATTGCGAG  
TTACAAGTTCTTATACTAAGTGGAACGATTTGGTTATGGCGCTAACCTC  
<H E L I H M V K A L G I G R N S

< ORF RF[4] C

510 520 530 540 550  
CCTTTAGTGGAAATAGTTTTCTTTAAAAATTTTACTCACATCTATTTGTTT  
GGAATCACCTTATCAAAACAAATTTTAAAAATGAGTGTAGATAACAAA  
<G K T S Y N E K T I K S V D I O K

< ORF RF[4] C

560 570 580 590 600  
TTCTTTGGTGGAAATCTGAATGATAAAGACTATACTGCCATTTTCTTAA  
AAGAAACCACCTTAAGACTTACTATTTCTGATATGACGGTAAAAAGAAAT  
<E K T S N O I I F V I S G N K K F

< ORF RF[4] C

610 620 630 640 650  
AAAAGGCTAACTGAATTTTCAGGATTTAATGATTGAAAGCAGCCTCAATG  
TTTTCCGATTGACTTAAAGTCTAAATTACTAAGCTTTTCGTCGGAGTTAC  
M>

< F A L O I E P N L S E P A A E I

< ORF RF[4] C

660 670 680 690 700  
GCATTATCACACAAGATAGAAAGTATGGTAATGAAATCAAGCAACTCCAT  
CGTAATAGTGTGTTCTATCTTTCATACCATTTAGTTTCGTTGAGGTA  
A L S H K I E S M V M K S S N S I>

< ORF RF[3]

< A N D C L I S L I T I F D L E E M

< ORF RF[4] C

710 720 730 740 750  
CTCAGGCGATTGTATTTTACTTGAGACTTCTACATTGACAGCAATCTTTT  
GAGTCCGTC AACATAAAATGAACTCTGAAGATGTAAGTGTCTGTTAGAAAA  
S G S C I L L E T S T L T A I F>

< ORF RF[3]

< E P L Q I K S S V E V N V A I K K

< ORF RF[4] C

760 770 780 790 800  
TATTCTGAGCTTCTAAGATTTTCTGCTGACAAGATACCCCTTGACAGCATCG  
ATAAGACTCGAAGATTCTAAAAACGACTGTTCTATGGGAAGTTCGTTAGC  
L F>

< N Q A E L I K A S L I G K U A D

< ORF RF[4] C

810 820 830 840 850  
TTTTGAATATTAGCTAGATGGCCGATATTATAACGGGTATCCTGCAATTG  
AAAACCTTATAATCGATCTACCGGCTATAATATTGCCCATAGGACGTTAAC

< \_\_\_\_\_ ORF RF[4] C

860                870                880                890                900  
ATGTCCTGTTTTTTCTAAGATTGTGATGGTAAATCTTTTCAATACTAGCTA  
TACAGGACAAAAAGATTCTAAACTACCATTTAGAAAAGTTATGATCGAT  
~~CH G T K E L I O R Y P K E T S A L~~

< \_\_\_\_\_ ORF RF[4] C

910                920                930                940                950  
AATCTTTTATTTTCAATGCCTAATCTCAGGCTAGTTAAATATTTACAGATAA  
TTAGAATAAAAAGTTACGGATTAGACTCCGATCATTTTATAAGTCTATT  
~~D K N E I G L R M S T L I E V L Y~~

< \_\_\_\_\_ ORF RF[4] C

960                970                980                990                1000  
TCATGGCGGAAACTTCGAATATCCTTGTAAGAGATTCTATTTGCTGACT  
AGTACCGCCTTTGAAGCTTATAGGAACATTTCTCTAAGATAAACCACTGA  
M A E T S N I L V K R F Y L L T>

~~D H R E S R I D K Y L S E L Q Q S~~

< \_\_\_\_\_ ORF RF[4] C

1010              1020              1030              1040              1050  
ATACTGGGTGATATTTTCAATCTGAGCTTCCTTTTGTCGCCATTATCTCAT  
TATGACCCACTATAAAGCTTAGACTCGAAGGAAAACACGGTAATAGAGTA  
I L G D I S N L S F L L C H Y L I

<~~V Q T I N R I Q A E K Q A M T E N~~

< \_\_\_\_\_ ORF RF[4] C

1060              1070              1080              1090              1100  
TTTGAACCTTTTGTTGGTATATTGGCTTAAAAATGAGATCAGAATCAAA  
AAACTTGGAAAAACAACCATATAACCGAATTTTTACTCTAGTCTTAGITT  
L N L L F G I L A>

>  
~~< Q V K Q K T Y Q S L F S L L L L~~

< \_\_\_\_\_ ORF RF[4] C

1110              1120              1130              1140              1150  
AATAAAATAAGATAGACAATAACGACAAATTTACGAAATTTTAAAGTCGG  
TTATTTTATTCTATCTGTTATTGCTGTTTAAATGCITTAATTTACGCC  
~~F L I L Y V P V V F K R F K L T P~~

< \_\_\_\_\_ ORF RF[4] C

1160              1170              1180              1190              1200  
TATCACATTATAACTCTCTATAACATACCTGTATTTAAAGGTAGT  
ATAGTGTAATATTGAGAGATATTGTATGTTATGGACATAATTTCCATCA  
~~I V N Y S E I V Y L V Q I L L Y Y~~

< \_\_\_\_\_ ORF RF[4] C

1210              1220              1230              1240              1250  
ATAGAAGCATAGTAATATTCATTGCAATCAAGCGTTTTTTGACCTTCATC  
TATCTTCGTATCATTATAAGTAACCTTAGTTCGCAAAAACTGGAAGTAG  
~~L L M T I N M E E L R K R V K M~~

< \_\_\_\_\_ ORF RF[4] C

1260 1270 1280 1290 1300  
TTGGTCAAACCTATCTTTAAGTCGACCAATATCAACATTGAACACACTGAG  
AACCAGTTTGATAGAAATTCAGCTGGTTATAGTTGTAACCTGTGTGACTC  
<K P L S D K L R G E D V N E F F S L  
< ORF RF[4] C

1310 1320 1330 1340 1350  
AAAGAGGTAACCTTAGCGCTATTCCTGCAAACCTCGATCATATAGGTGGTTA  
TTCTCCATTGAATCGCGATAAGGACGTTTGAGCTAGTATATCCACCAAT  
<F L Y S L A I G A F E T M V T T I  
< ORF RF[4] C

1360 1370 1380 1390 1400  
TAATGCTACTGCCCATTAACAATTCCTTGAGTTCCATCCAAGATAAAGAAT  
ATTACGATGACGGGTAATGTTAAGGAACTCAAGGTAGGTTCTATTCTTA  
<I S S G M V I G Q T G D L I F F  
< ORF RF[4] C  
<L E K L E M W S L S Y  
<

1410 1420 1430 1440 1450  
ATGATTGCCCGCCTAAACAAGTCAGAACTGGCAACAGGCAGCAGACCATA  
TACTAACCGCGCGATTGTTTCAGTCTTGACCGTTGTCCGTCGTCGTGTAT  
<I I A P R E F L D S S A V P D L L Q Y  
< ORF RF[4] C  
<S Q G G L C T L V P L L C C V M  
<

1460 1470 1480 1490 1500  
AAATATATTTAGAGAAAGACTATTCTGTCTATTAAGATAAATTGATAAAG  
TTTATATAAATCTCTTTCTGATAAGACAGATAATTCTATTTAACTATTTTC  
<F I N L L S L S N O F N L Y T L S L A  
< ORF RF[4] C

1510 1520 1530 1540 1550  
CTATAAAATAAAGAGGCTCTGCGAGATAAAACAGGTTTACGTTCCACATC  
GATATTTTATTTCTCCGAGACGTCCTATTTTGTCCAAATGCAAGTGGTAG  
<I F F L P E A P V F L N V N V M  
< ORF RF[4] C

1560 1570 1580 1590 1600  
GTAACAGCAATCATTATCAGAAAATTGCTTATCGAAAAAGAGTTAATTC  
CATTGTCGTTAGTAATAGTCTTTTAACGAATAGCTTTTTCTCAATTAAG  
<T V A I N I L E N S I S F L T P E  
< ORF RF[4] C

1610 1620 1630 1640 1650  
CTTTTTCGATAAAGTGACATTACTTACCTTAGAAAATAGAAACAAGAGAA  
GAAAAGCTATTTCACTGTAATGAATGGAATCTTTTATCTTTGTTCTTT  
<K K S L T V N S V K S F L F L L F  
< ORF RF[4] C

1660 1670 1680 1690 1700  
ATAGAACGGTTAGATAAGTTAATAAACCATTTGAAAGTATCATTAAAGGCT  
TATCTTGCCAATCTATTCAATTATTTGCTAAACTTTCATAGTAATTCGGA  
<L V T L Y T L L G N S L I N L A  
< ORF RF[4] C

09833017 0400

1710 1720 1730 1740 1750  
TCATTCAATTTGCTCTCTCTTTGATCAGCAATCACAGCTCTCAGTTTTGTT  
AGTAAGTAAACCGAGAGGAACTAGTCGTTAGTGTGAGAGTCAAAACAA

<E N M

<K A R R Q D A I V A R L K T  
< ORF RF[5] C

1760 1770 1780 1790 1800  
AACTTAGTTCGTGAAATAAGACAAGACTTATTATTTTGGAAAATAGGCCAA  
TTGAATCAAGCACTTTATTCTGTTCTGAATAATAAGCTTTTATCCGCTT  
<L K I R S L L C S K N N R F Y A T

< ORF RF[5] C

1810 1820 1830 1840 1850  
GCGTTTTTCCCATCAATACGGGTAATATTGGCAGGATTGACAATAAAG  
CGCAAAAAGCGCTAGTTATGCCCATTTATAACCGTCCTAACTGTTATTTTC  
<R K K R D L R T L N A P N V L P S

< ORF RF[5] C

1860 1870 1880 1890 1900  
AGCGATGGCACTGAAAAAGTCTCTTATCCATTTTAACAATGTCAGTCATA  
TCGCTACCGTGACTTTTTTCAGAGAATAGGTAAAATTGTTACAGTCAGTAT  
M A L K K S L I H F N N V S H>

ORF RF[1]  
<R H C Q P L R K D M K V I D T M  
< ORF RF[5] C

<V

<\_

1910 1920 1930 1940 1950  
CTGCCGTAGAATTCAATCCGTTTCATCATAAGTATAAAGGCAGAGCTTATG  
GACGGCATCTTAAGTTAGGCAAGTAGTATTTCATATTTCCGTCTCGAATAC  
T A V E F N P F I I S I K A E L M>

ORF RF[1]  
<S G Y F E I R E D Y T Y L C L R H

< ORF RF[5] C  
<A T S N L G N M M L I F A S S I  
< ORF RF[6] C

1960 1970 1980 1990 2000  
GGCTGTGAAGATGTTTCAAATACAGAATCTCAGCAAAAGGGACCTGAA  
CCGACAACTTCTACAAAGTTTATGTCTTAGAGTCGTTTTCCCTGGACTT  
G C>

<A T S S T E F Y L I E A F P V Q F  
< ORF RF[5] C

<P Q Q L H K L I C F R L L L S R F  
< ORF RF[6] C

2010 2020 2030 2040 2050  
ACTGAGTTTCAGATGAATGGAAAATAAAAAGTTCTCTGATTGACCATTC  
TGAATCAAAGTCTACTTACCTTTTATTTTCAAGGAGACTAACTGGTAAG  
<Q T E S S H F I F L E E S Q G N

< ORF RF[5] C  
<S L K L H I S F L F N R Q N V M R

09333017 041001

< ORF RF[6] C

2060 2070 2080 2090 2100  
TTCTGGCTGTTTTCCATAGCATAATACAGCGCTGATTCAATGCGGTGGGA  
AAGACCGACAAAAGGTATCGTATTATGTCCGCGACTAAGTTACGCCACCCT  
<K Q S N E M A Y Y L A S E I E E S

< ORF RF[5] C

<R A T K W L M

< ORF RF[6] C

2110 2120 2130 2140 2150  
GAACTCCTCAGGATTCAAAGATTTATCAATAAAATCCAAAGCAGATACCT  
CTTGAGGAGTCCTAAGTTTCTAAATAGTTATTTAGGTTTCGTCTATGGA  
<F E E P N I S K D I F D E A S V Q

< ORF RF[5] C

2160 2170 2180 2190 2200  
GATACTGAAAAGTGAGGGGCATAAACTCAGAATGTGTGCTGACAAAGACA  
CTATGACTTTTCACTCCCCGTATTTGAGTCTTACACAGCACTGTTTCTGT  
M C R D K D>

<Y Q F T L E M F E S H I T V F V

< ORF RF[5] C

2210 2220 2230 2240 2250  
ATAACTGCCTAGGATTATGCTGTCTAATCTGATTGGCTACTTCCAGTCC  
TATTGACGTGATCCTAATACGACAGATTAGACTAACCGATGAAGGTCAGG  
N N C T R I M L S N L I G Y F Q S>

<I V A S D N H O R I O N A V E L G

< ORF RF[5] C

2260 2270 2280 2290 2300  
TTTCTTTTCTCTTTTGTATTTCAATATCCAAAAGAAAATCTGGTGAT  
AAAGAAAAGGAGAAAAAATAAGTTATAGGTTTCTTTTAGACCACTA  
F L F L F F D F N I Q K E N L V I>

<K K E E K K I E I D L F F I Q H N

< ORF RF[5] C

2310 2320 2330 2340 2350  
TGCCCTTTTCAGGGATAGCGTCAATAAGTTGTTGTTGGTTTTCCAAAATA  
ACGGGAAAAGTCCCTATCGCAGTTATTCAACAACACCAAAGGTTTTTAT  
A L F R D S V N K L L W F S K N>

<G K E P I A D I L Q Q P K G F I

< ORF RF[5] C

2360 2370 2380 2390 2400  
GTCAATTCTTTATAAGACCAATTTTTCTTTTCATGATAGCTGCAATCGT  
CAGTTAAGAAATATTCTGGTTAAAAAAGAAAGTACTATCGACGTTACCA  
S Q F F I R P I F F F H D S C N G>

M I A A M V>

<T L E K Y S W N K E K M L A A I T

< ORF RF[5] C

09833017 "041001

Express Mail No. EL818623952US  
Inventor(s): Cvitkovitch et al  
Atty Dkt No.: 1889-00401  
Contact: C.M. Mintz (713)238-8000  
Sheet 15 of 19

2410 2420 2430 2440 2450  
GGTTTCAGACGTCCTTGTTGTAAAAATCATCTTCCAATACAAAAATAG  
CCAAAGTTCTGCAGGAACAACATTTTTTAGTAGAAGGTTATGTTTTTATC  
G F K T S L L>

>  
V S R R P C C K K S S S N T K I>

<T E G R G Q Q L P D D E L V E T S  
< ORF RF[5] C

2460 2470 2480 2490 2500  
AAATCATTATTTCTCCITTAATCTTCTATTAGGTTAGCTGATTAACACT  
TTTAGTAATAAAGAGGAAATTAGAAGATAAATCCAATCGACTAATTGTGA  
E I I I S P L I F Y L G>

<I M

2510 2520 2530 2540 2550  
ATACACAGAAAAGGTATAAACGATATCACTCAATAAAATCTACTAACTT  
TATGTGTCTTTTCCATATTTTGCTATAGTGAGTTATTTTAGATGATTGAA

AATAACC  
TTATTGG

093301-041001

Figure 10

The comX nucleotide sequence, amino acid sequence, and its local region with 100bp included both upstream and downstream (promoter is upstream).

> *S. mutans* comX gene

ATGGAAGAAGATTTTGAAATTGTTTTTAATAAGGTAAAGCCAATTGTATGGAAATTAAG  
CCGTTATTACTTTTATTAATAATGTGGACTCGTGAAGATTGGCAACAAGAGGGAATGTTGA  
TTTTCACCAATTATTAAGGGAACATCCAGAATTAGAAGAGGATGATACAAAATTGTAT  
ATCTATTTTAAGACACGTTTTTCTAATTACATTAAAGATGTTTTGCGTCAGCAAGAAAG  
TCAGAAACGTCGTTTTAATAGAATGTCTTATGAAGAAGTCGGTGAGATTGAACACTGTT  
TGTCAGTGGCGGTATGCAATTGGATGAATATATTTTATTCGTGATAGTTTGCTTGCA  
TATAACAAGGTCTGAGTACTGAAAAGCAAGAGCTGTTTGAGCGCTTGGTAGCAGGAGA  
GCACTTTTTGGGAAGGCAAAGTATGCTGAAAGATTTACGTAAAAAATTAAGTGATTTTA  
AGGAAAAA

> *S. mutans* ComX protein

MEEDFEIVFNKVKPIVWKLRSYYFIKMWTRWDQQEGMLILHQLLREHPELEEDDTKLY  
IYFKTRFSNYIKDVLRQESQKRRFNRMSEYEEVGEIEHCLSSGGMQLDEYILFRDSL  
LLAYKQGLSTEKQELFERLVAGEHPLGRQSMKDLRKKLSDFKEK

> *S. mutans* comX gene local region

GTAAATAAAACAGCCAGTTAAGATGGGACATTTATGTCCTGTTCTTAAAGTCTTTTTTCG  
TTTTATAATAATTTTATTATAAAAGGAGGTCATCGTAATAGATGCAAGAAGATTTTGAA  
ATTGTTTTTAATAAGGTAAAGCCAATTGTATGGAAATTAAGCCGTTATTACTTTATTAA  
AATGTGGACTCGTGAAGATTGGCAACAAGAGGGAATGTTGATTTTGCACCAATTATTAA  
GGGAACATCCAGAATTAGAAGAGGATGATACAAAATTGTATATCTATTTTAAGACACGT  
TTTTCTAATTACATTAAAGATGTTTTGCGTCAGCAAGAAAGTCAGAAACGTCGTTTTAA  
TAGAATGTCTTATGAAGAAGTCGGTGAGATTGAACACTGTTTGTCAGTGGCGGTATGC  
AATTGGATGAATATATATTTTATTTTCGTGATAGTTTGCTTGCATATAACAAGGTCTGAGT  
ACTGAAAAGCAAGAGCTGTTTGAGCGCTTGGTAGCAGCAGAGCACTTTTTGGGAAGGCA  
AAGTATGCTGAAAGATTTACGTAAAAAATTAAGTGATTTTAAGGAAAAATAGTTAAAAA  
GGGAAAGAATGGAACATGTGATTGTACCATTCTTTTGGTTGAAAATTAAGAAAAGTTA  
TTATAAATTATTGGTTTAACATGCCATATTA

FIG 10



Figure 11.

The comA and comB nucleotide and amino acid sequences. ComA and ComB are the components of the CSP exporter.

> *S. mutans* comA gene

ATGAAACAAGTTATTTATGTTGTTTTAATCGTCATAGCCGTTAACATTCTCTTAGAGAT  
TATCAAAAGAGTAACAAAAGGGGAGGGACAGTTTCGTCATCTAATCCTTTACCAGATG  
GGCAGTCTAAGTTGTTTTGGCGCAGACATTATAAGCTAGTACCTCAGATTGATACCAGA  
GACTGTGGGCCCGCAGTGTGTCATCTGTTGCAAAGCATTACGGATCTAATTACTCTAT  
CGCTTATCTGCGGGAACCTCTCAAAGACTAAACAAGCAGGGAACAACAGCTCTTGGCATTG  
TTGAAGCTGCTAAAAAGTTAGGCTTTGAAACACGCTCTATCAAGGCGGATATGACGCTT  
TTTGATTATAATGATTTGACCTATCCTTTTATCGTCCATGTGATTAAAGGAAAACGTCT  
GCAGCATTATTATGTGCTCTATGGCAGCCAGAATAATCAGCTGATTATTGGAGATCCTG  
ATCCTTCAGTTAAGGTGACTAGGATGAGTAAGGAACGCTTCAATCAGAGTGGACAGGC  
CTTGCAATTTTCTAGCTCCTCAGCCTAACTATAAGCCTCATAAAGGTGAAAAAATGG  
TTTGTCTAATTTCTCCCGTTGATCTTTAAGCAGAAAGCTTTGATGACTTATATTATCA  
TAGCTAGCTTGATTGTGACGCTCATTGATATTGTGCGATCATACTATCTCCAAGGAATA  
TTGGACGAGTACATTCTGATCAGCTGATTTCAACTTTAGGAATGATTACGATTGGTCT  
GATAATAACCTATATTATCCAGCAGGTCTGCTTTTGCAAAGAATACCTCTTGCCCG  
TACTCAGTTTGCGTTTAGTCATTGATGTTATCCTGTCTTATATCAAACATATTTTACG  
CTTCCTATGCTTTCTTTGCGACAAGGCGAACAGGAGAAATCACGTCTCGTTTTACAGA  
TGCCAATCAGATTATTGATGCTGTAGCGTCAACCATCTTTTCAATCTTTTTAGATATGA  
CTATGGTAATTTTGGTTGGTGGGTTTTGTTGGCGCAAAACAATAACCTTTTCTTTCTA  
ACCTTGCTCTCCATTCCGATTATGCCATCATTATTTTGGCTTTCTTGAAACCTTTGA  
GAAATGAATCACGAAGTGATGGAAGCAATGCTGTGGTAAGTTCTTCTATCATTGAAG  
ATATCAATGGGATGGAACCATTAATCACTCACAAAGTGAGTCCGCTCGTTATCAAAAC  
ATTGATAGTGAATTTGTTGATTATTTGGAGAAAACTTTAAGCTACACAAGTATAGTGC  
CATTCAAACCGCATTAAAAAGCGGTGCTAAGCTTATCCTCAATGTGTGTCATTCTCTGGT  
ATGGCTCTCGTCTAGTTATGGATAATAAAATCTCAGTTGGTCAGCTTATCACCTTTAAT  
GCTTTGCTGTCTTATTTCTCAAATCCAATGAAAATATTATCAATCTGCAATCCAACT  
GCAGTCAGCTCGCGTTGCCAATACACGTCTTAATGAGGTCTATCTTGTCGAATCTGAAT  
TTGAAAAGACGGCGATTTATCAGAAAATAGCTTTTTAGATGGTGATATTTGTTTTGAA  
AATCTTCTTATAAATATGGATTTGGGCGAGATACCTTATCAGATATTAATTTATCAAT  
CAAAAAGGCTCCAAGGTCAGTCTAGTTGGAGCCAGTGGTTCTGGTAAACAACCTTTGG  
CTAAACTGATTGTCAATTTCTACGAGCCTAACAAGGGGATTGTTTGAATCAATGGCAAT  
GATTTAAAAGTTATTGATAAGACAGCTTTGCGGCGGCATATTAGCTATTTGCCGCAACA  
GGCCTATGTTTTTAGTGGCTCTATTATGGATAATCTCGTTTTAGGAGCTAAAGAAGGAA  
CGAGTCAGGAAGACATTATTCGTGCTTGTGAAATTGCTGAAATCCGCTCGGACATTGAA  
CAAATGCCTCAGGGCTATCAGACAGAGTTATCAGATGCTGCGGTATTTCTGGCGGTCA  
AAAACAGCGGATTGCTTTAGCTAGGGCCTTATTAACACAGGCACCGGTTTTGATTCTGG  
ATGAAGCCACCAGCAGTCTTGATATTTTGACAGAAAAGAAAATTATCAGCAATCTCTTA  
CAGATGACGGAGAAAACAATAATTTTGTGCCCACCGCTTAAGCATTTACAGCGTAC  
TGACGAAGTCATTGTCATGGATCAGGGAATAATTGTTGAACAAGGCACTCATAAGGAAC  
TTTTAGCTAAGCAAGGTTTCTATTATAACCTGTTTAAT

> *S. mutans* ComA protein

09833047-041001

MKQVIYVVLIVIAVNILLEIIKRVTKRGGTVSSSNPLPDGQSKLFWRRHYKLVPQIDTR  
DCGPAVLASVAKHYGSNYSIAYLRELSKTNKQGTALGIVEAAKKLGFETRSIKADMTL  
FDYNDLTYPFIVHVIKGRLOHYYVVGYSQNNQLIIGDPDPSVKVTRMSKERFQSEWTG  
LAIFLAPQPNYKPHKGEKNGLSNFFPLIFKQKALMTYIIIIASLIVTLIDIVGSYYLQGI  
LDEYIPDQLISTLGMITIGLIITYIIQQVMAFAKEYLLAVLSLRLVIDVILSYIKHIFT  
LPMSFFPATRRRTGEITSRFTDANQIIDAVASTIFSIFLDMTMVILVGGVLLAQNMMNLFFL  
TLLSIPYAIIIIFAPLKPFEKMNHEVMESNAVVSIIIEDINGMETIKSLTSESARYQN  
IDSEFVDYLEKNFKLHKYSAIQTALKSGAKLILNVVILWYGSRLVMDNKISVGQLITFN  
ALLSYFSNPNIENIINLQSKLQSARVANTRLNEVYLVESEFEKDGDLSENSFLDGDISFE  
NLSYKYGFGRDTLSDINLSIKKGSKVS LVGASGSGKTTLAKLIVNFYEPNKGIVRINGN  
DLKVIDKTALRRHISYLPQQAYVFSGSIMDNVLGAKEGTSQEDIIRACEIAEIRSDIE  
QMPQGYQTELSDGAGISGGQKQRIALARALLTQAPVLILDEATSSLDILTEKKIISNLL  
QMTKTIIFVAHRLSISQRTDEVIVMDQKIVEQGTHKELLAKQGFYNNLFN

> S. mutans comB gene

ATGGATCCTAAATTTTACAAAGTGCAGAATTTTATAGGAGACGCTATCATAATTTTGC  
GACACTATTAATTGTTCTTTGGTCTGCTTGATTATCTTCTTGGTCATATTCCTTTGTT  
TTGCTAAAAAAGAAATTACAGTGATTTCTACTGGTGAAGTTGCACCAACAAAGGTTGTA  
GATGTTATCCAATCTTACAGTGACAGTTCAATCATTAAAAATAATTTAGATAATAATGC  
AGCTGTTGAGAAGGGAGACGTTTTAATTGAATATTAGAAAATGCCAGTCCAAACCGTC  
AGACTGAACAAAAGAATATTATAAAAGAAAGACAAAAACGAGAAGAGAAGGAAAAGAAA  
AAACACCAAAAGAGCAAGAAAAAGAAAGTCTAAGAGCAAGAAAGCTTCCAAAGATAA  
GAAAAAGAAATCGAAAGACAAGGAAAGCAGCTCTGACGATGAAAAAGAGACAAAAAGG  
TTTCGATTTTTGCTTCAGAAGATGGTATTATTATACCAATCCCAATATGATGGTGCC  
AATATTATTCGAAGCAAACCGAGATTGCTCAAATCTATCCTGATATTCAAAAACAAG  
AAAAGTGTTAATCACCTATTATGCTTCTTCTGATGATGTTGTTTCTATGAAAAAGGGGC  
AAACCGCTCGTCTTTCCTTGAAAAAAGGGAAATGACAAGGTTGTTATTGAAGGAAAA  
ATTAACAATGTGCTTCATCAGCAACTACTACTAAAAAAGGAAATCTCTTTAAGGTTAC  
TGCCAAAGTAAAGGTTTCTAAGAAAAATAGCAAACCTCATCAAGTATGGTATGACAGGCA  
AGACAGTCACTGTCAATTGATAAAAAGACTTATTTTGATTATTTCAAAGATAAATTACTG  
CATAAAATGGATAAT

> S. mutans ComB protein

MDPKFLQSAEFYRRRYHNFATLLIVPLVCLIIIFLVIFLCFAKKEITVISTGEVAPTKVV  
DVIQSYSDSSIIKNLNDNNAAVEKGDVLI EYSENASPNRQTEQKNIIKERQKREEKEKK  
KHQSKSKKKKSKSKKASKDKKKKSKDKESSDDENETKKVSIFASEDGIHTNPKYDGA  
NIIPKQTEIAQIYPDIQKTRKVLITYYASSDDVVSMMKKGQTARLSLEKKGNDKVVIEGK  
INNVASSATTTKGNLFKVTAKVKVSKKNSKLIKYGMTGKTVTVIDKKTYFDYFKDKLL  
HKMDN

041001 09533017 000000

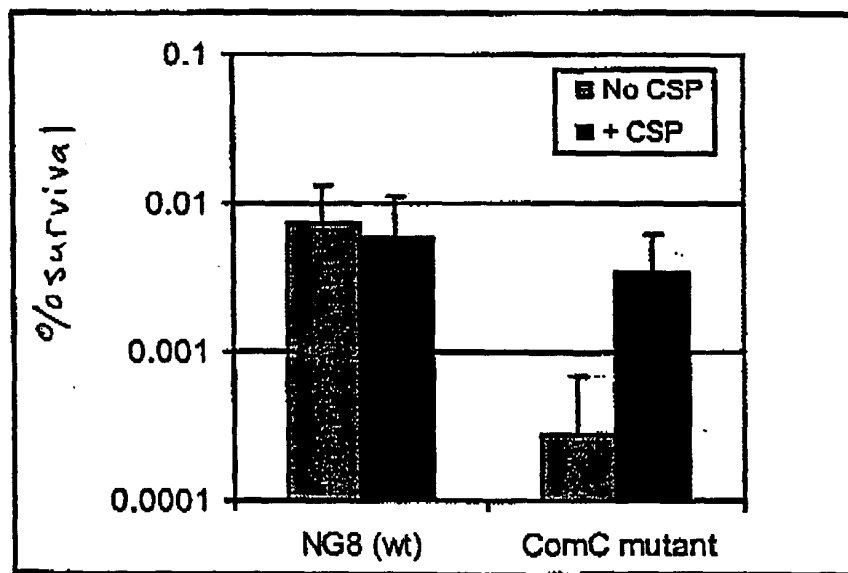


Figure 12